

ORION-CMR: On-scanner Reporting with Integrated Foundation Model for End-to-End Cardiac MRI Analysis and Interpretation

Omer Burak Demirel^{1,2*}, Kelly K. Horst^{2*}, Alessio Perazzolo², Elisa Bruno², Kenan Kaya², Rongzhen Ouyang², Enas Ahmed², Jouke Smink¹, Spencer L. Waddle^{1,2}, Tzu Cheng Chao², Dinghui Wang², Steve G. Langer², Timothy L. Kline², Panagiotis Korfiatis², Jacinta Browne², Ivana Isgum², and Tim Leiner²

¹ MR Clinical Science, Philips North America, MN, USA

{burak.demirel,spencer.waddle,jouke.smink}@philips.com

² Department of Radiology, Mayo Clinic Rochester, Rochester, MN, USA

{Horst.Kelly,Perazzolo.Alessio,Bruno.Elisa,kaya.kenan,

Ouyang.Rongzhen,chao.tzucheng,wang.dinghui,Ahmed.Enas,

browne.jacinta,Langer.Steve,Kline.Timothy,

Korfiatis.Panagiotis,igum.ivana,Leiner.Tim}@mayo.edu

*Equal contribution

Abstract. Cardiovascular magnetic resonance (CMR) provides comprehensive cardiac assessment but remains underutilized because of the complexity of acquisition, post-processing, and interpretation. Existing artificial intelligence (AI) methods address isolated tasks, limiting clinical integration. We present ORION-CMR (On-scanner Reporting with Integrated Foundation Model), the first clinically evaluated scanner-native end-to-end CMR foundation model. Pretrained on 12,896,733 CMR images from 9,258 studies, ORION-CMR performs sequence classification, ventricular function assessment, late gadolinium enhancement (LGE) detection, binary and multiclass disease classification, and local large language model-based report generation in approximately 90 seconds. The framework was evaluated on public benchmarks and clinically validated in a multi-vendor cohort of 68 subjects with normal examinations, congenital heart disease, dilated cardiomyopathy, and myocardial infarction. ORION-CMR outperformed supervised baselines and the previously published CMR foundation model (CMR-FM), achieving state-of-the-art performance for LGE classification and scar segmentation. Clinical evaluation achieved an AUC of 0.96 for normal-versus-abnormal classification and 0.88 for multiclass disease classification, while generated reports demonstrated 81.4% agreement with expert interpretation. These results demonstrate the feasibility of real-time scanner-native AI-assisted CMR analysis and automated report generation.

Keywords: foundation model · cardiac magnetic resonance · classification · segmentation · structured reporting · on-scanner AI.

1 Introduction

Cardiovascular magnetic resonance (CMR) is the reference standard for quantitative assessment of cardiac structure and function, including ventricular volumes and ejection fraction, yet remains underutilized in routine clinical practice [1, 12, 22]. The complexity of CMR acquisition, interpretation, and post-processing, together with the need for specialized expertise and software, limits accessibility compared with other imaging modalities [8]. A recent US national analysis reported fewer than 1,000 physicians interpreting CMR studies, with nearly 70 million Americans living more than 50 miles from a CMR service location [9]. Similarly, increasing demand for CMR has raised concerns about broader adoption worldwide [6, 19].

CMR provides comprehensive characterization of cardiac structure, function, and tissue composition without ionizing radiation, enabling longitudinal disease assessment [12, 3, 13]. Late gadolinium enhancement (LGE) imaging identifies myocardial tissue damage with important diagnostic and prognostic value [1, 12]. However, comprehensive CMR interpretation requires integrating multiple image sequences, quantitative measurements, and clinical context across fragmented software workflows [8, 3], increasing interpretation time, inter-reader variability, and limiting clinical scalability.

Deep learning has substantially improved automated CMR analysis, yet most methods remain designed for a single specific task such as segmentation, classification, or tissue characterization. Foundation models provide unified representations that support multiple downstream tasks from a single encoder [22, 14, 29]. This paradigm has improved data efficiency and downstream performance across CT, ultrasound, and MRI [14, 18, 10]. Recent CMR studies have adopted self-supervised Vision Transformers (ViTs) for segmentation and disease classification [29, 14]. However, these approaches still focus on individual downstream tasks rather than unified, clinically deployable workflows.

In this work, we present ORION-CMR (On-scanner Reporting with Integrated fOundatioN Model), an MRI scanner-native end-to-end CMR interpretation framework. A single pretrained encoder supports sequence classification, ventricular function assessment, LGE classification, binary and multiclass disease classification, and local large language model (LLM)-based report generation. The complete pipeline is implemented directly on the MRI scanner and clinically evaluated, including the generated reports.

2 Methods

2.1 ORION-CMR Framework

ORION-CMR is a self-supervised CMR foundation model for end-to-end scanner-native CMR interpretation. A shared ViT encoder supports sequence classification, cine short-axis (SAX) segmentation, LGE detection, and disease classification. Quantitative measurements and categorical findings are consolidated into

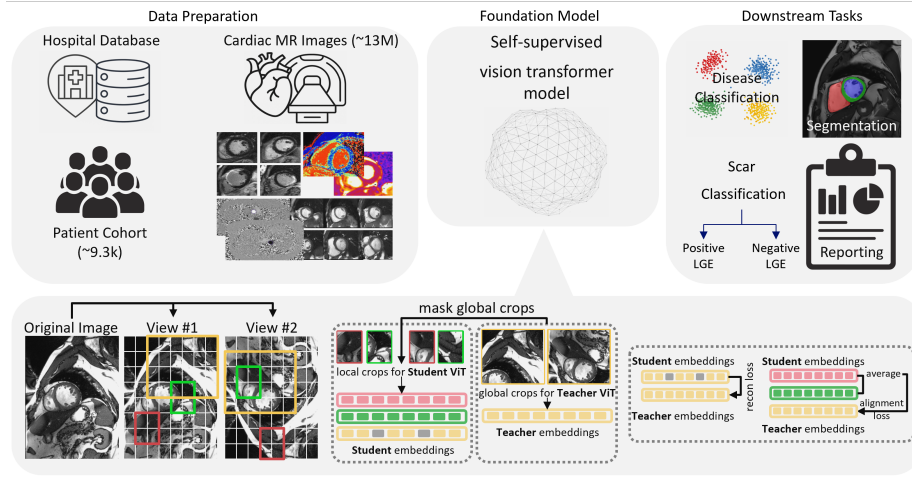


Fig. 1. Overview of the ORION-CMR (pretrained on 9,258 CMR studies) framework for scanner-native end-to-end CMR analysis and automated report generation.

a structured representation that is processed by a locally deployed LLM to generate a clinical report. The complete workflow executes on the MRI scanner in real time. A schematic overview is shown in Fig. 1.

2.2 Patient and Data Characteristics

ORION-CMR was pretrained using 12,896,733 CMR image slices from 9,258 multi-vendor CMR studies spanning localizer, cine, LGE, T1/T2/T2* mapping, perfusion, and phase-contrast flow acquisitions. The mean patient age was 55.5 ± 17.2 years and 45.9% (4,247/9,258) of patients were female. The dataset comprised retrospectively collected consecutive CMR studies acquired at Mayo Clinic between 2017 and 2025 on GE Healthcare and Siemens Healthineers MRIs.

The Automatic Cardiac Diagnosis Challenge (ACDC) [2] and Automatic Evaluation of Myocardial Infarction from Delayed-Enhancement Cardiac MRI (EMIDEC) [16] datasets were used for public benchmark evaluation. ACDC cine-SAX images were used for disease classification and ventricular segmentation, while EMIDEC PSIR-LGE images were used for infarct classification and scar segmentation. Clinical validation used an in-house multi-vendor cohort of 426 subjects (168 normal, 66 congenital heart disease, 80 dilated cardiomyopathy, and 112 myocardial infarction), partitioned into a 358-subject development cohort for LGE and multiclass disease classification and an independent 68-subject hold-out testing cohort (25 normal, 7 congenital heart disease, 15 dilated cardiomyopathy, and 21 myocardial infarction) for cine-SAX segmentation, LGE classification, binary and multiclass disease classification, and automated report generation.

2.3 Self-Supervised Foundation Model Training

Pretraining employed a Self-Distillation with NO labels (DINO)-style teacher-student framework [5, 20] with a ViT-Base (ViT-B) backbone and no manual labels. Two variants with patch sizes of 16×16 and 8×8 pixels (ViT-B16 and ViT-B8, respectively) were investigated to assess whether finer patch representations improve characterization of small anatomical structures and tissue abnormalities in CMR. The student and teacher networks shared the same architecture, with teacher weights updated using an exponential moving average of the student’s.

All CMR images were loaded from a curated study manifest, converted to standardized three-channel inputs, and augmented using a multi-crop strategy with two global and six local views per image. Global views were resized to 224×224 pixels, whereas local views were randomly cropped to 96×96 pixels. Augmentation comprised percentile-based intensity clipping and normalization, random resized cropping, horizontal flipping, rotation, brightness and contrast jittering, Gaussian blurring, elastic deformation, and additive Gaussian noise. Images were normalized with channel-wise mean and standard deviation values of 0.5. Optimization used AdamW with a base learning rate of 5×10^{-5} , cosine learning-rate decay, and a 10-epoch warmup. Teacher momentum increased from 0.996 to 1.0 following a cosine schedule, while teacher temperature increased from 0.04 to 0.07 during warmup. During training, the teacher processed the two global views and the student all eight augmented views. The resulting encoder served as the shared backbone for all downstream tasks.

2.4 Downstream Tasks

The pretrained ORION-CMR encoder was transferred to multiple downstream tasks where the foundation model’s encoder remained fixed and lightweight task-specific models were trained on top of the learned representations.

Sequence Classification In-house images were categorized into eleven common CMR sequence types by four expert radiologists, including localizer, cine 2-3-4ch-SAX, LGE, flow, perfusion, T1/T2/T2* mapping. The pretrained ViT encoder used concatenated feature embeddings from the final two transformer blocks, yielding a 1536-dimensional representation followed by a multilayer perceptron (MLP) classifier trained using cross-entropy loss and AdamW optimization. Performance was evaluated at the subject level. The predicted sequence labels were subsequently used to automatically identify the required image series for downstream tasks.

Segmentation Cine-SAX and LGE-SAX were performed where input images were first resized to 256×256 pixels, intensity-normalized, and converted to three-channel images. Patch tokens from the final four transformer blocks were concatenated, reshaped into spatial feature maps, and processed by a lightweight convolutional decoder to generate full-resolution segmentations. The decoder was

trained for 50 epochs using AdamW (learning rate 2×10^{-4} , weight decay 10^{-4} , batch size 8) with a combined class-weighted cross-entropy and soft Dice loss.

For the ACDC benchmark, cine-SAX segmentation targeted the right ventricle (RV), myocardium (Myo), and left ventricle (LV), whereas EMIDEC targeted Myo and scar. As in-house segmentations were unavailable, the ACDC-trained model was applied to the clinical cine-SAX cohort without additional training. The resulting segmentations were used to derive LV and RV parameters, which were incorporated into automated report generation.

LGE and Disease Classification CLS-token embeddings from the pretrained ORION-CMR encoder were extracted, mean-pooled at the subject level, L2-normalized, and classified using calibrated linear support vector machine (SVM) classifiers within a linear-probing framework. Hyperparameters and decision thresholds were optimized exclusively on the validation cohort.

For ACDC, cine-SAX-based disease classification used classification token embeddings from the final four transformer blocks (3072-dimensional). LGE-SAX classification used embeddings from the final two transformer blocks (1536-dimensional) for binary infarct classification for EMIDEC and the in-house cohort. Unlike EMIDEC, which contains only phase-sensitive inversion recovery (PSIR) images, the in-house cohort included PSIR, inversion recovery (IR), and motion-corrected (MOCO) LGE acquisitions, requiring a dedicated LGE-SAX classifier. Because CMR sequence provides complementary anatomical and tissue characterization information, disease classification used cine 2ch, 4ch, SAX, and LGE-SAX examinations. Classification token embeddings from the final two transformer blocks were concatenated into a 4×1536 -dimensional subject-level representation for multiclass disease classification (normal, congenital heart disease, dilated cardiomyopathy, and myocardial infarction), from which binary normal-versus-abnormal classification was derived.

Report Generation Outputs from ventricular segmentation, LGE classification, and disease classification were consolidated into a structured study representation containing quantitative measurements and AI model predictions. Ventricular volumes, ejection fractions, and indexed left ventricular mass were interpreted using age-, sex-, and ethnicity-specific CMR reference values from the Healthy Hearts Consortium (HHC) 2024 framework to derive standardized clinical findings, including ventricular size, systolic function, and myocardial mass classifications [21, 15]. The HHC framework was selected as the primary reference standard because it provides a large contemporary multi-ethnic CMR reference dataset spanning the adult age spectrum, while additional normative CMR reference studies informed pipeline development and verification [28].

The resulting structured findings, together with quantitative ventricular measurements, LGE classifier outputs, and disease classifier predictions, were assembled into a deterministic internal representation of the examination. Automated ventricular measurements were obtained using deep learning-based segmentation, while myocardial tissue characterization was derived from dedicated LGE

classifiers [24, 26]. This verified structured representation was then provided to a locally deployed Qwen2.5-14B-Instruct LLM running through Ollama, which synthesized a complete radiologist-style CMR report from the protected findings [25].

2.5 Evaluation

For evaluation, the LLM-generated reports were compared directly with the original clinical reports authored by expert radiologists. This design enabled assessment of whether an LLM could synthesize verified quantitative measurements and structured findings into clinically coherent reports that closely resembled expert radiologist reports while maintaining factual consistency and minimizing unsupported inferences or hallucinations [4, 27].

For image analysis tasks, ORION-CMR was compared with supervised baseline models, the previously published CMR cardiac foundation model (CMR-FM) [14], and reported state-of-the-art (SoTA) results on the public benchmark datasets when available [2, 17]. Supervised baselines consist of a ResNet-50 classifier initialized with ImageNet weights for sequence classification [11], U-Net architectures for segmentation [23], and ResNet-18 classifiers for disease and LGE classification [11]. All baseline models were trained and evaluated using identical dataset partitions, preprocessing, and evaluation metrics.

Segmentation performance was assessed using the Dice index (D) and Hausdorff Distance (d_H (mm)). Sequence, LGE, and disease classification performance were evaluated using area under the curve (AUC), accuracy, sensitivity, and specificity. Generated reports were independently evaluated against original reports by three expert readers using a three-category agreement scale: (1) concordant (no or minimal likelihood of changing treatment decisions), (2) intermediate (possible impact on treatment decisions), and (3) discordant (high likelihood of different treatment decisions). Inter-reader agreement was assessed using Cohen’s κ for pairwise comparisons. Evaluation focused on the report impression, which contains the key clinical message. Confidence intervals and statistical comparisons were performed using bootstrap resampling and paired significance testing where appropriate ($P < 0.05$ was considered statistically significant).

2.6 On-Scanner Implementation

The complete ORION-CMR pipeline was deployed on a Philips 1.5T Ambition X MRI scanner equipped with an NVIDIA RTX A6000 GPU [7]. Although the clinical validation cohort comprised retrospectively acquired multi-vendor examinations, the identical scanner-native workflow was applied to all studies. As each image series became available, ORION-CMR first performed sequence classification to identify the acquisition type and automatically trigger the corresponding downstream task (e.g., segmentation or LGE/disease classification). Quantitative measurements and categorical findings were progressively accumulated into a structured representation throughout the examination. Following completion

of all required sequences, the locally deployed LLM generated a clinical report, which was exported as a new DICOM series without external network communication, preserving patient privacy.

Table 1. Disease classification performance on ACDC cine-SAX and infarct classification performance on EMIDEC LGE-SAX. Best results are shown in bold.

Method	ACDC Disease (cine-SAX)						EMIDEC LGE (LGE-SAX)
	Overall	NORM	DCM	HCM	RV	MINF	Overall
ResNet-18 [11]	0.480	0.300	0.600	0.200	0.600	0.700	0.667
SoTA [17]	0.960	1.00	0.899	1.000	0.899	1.000	0.920
CMR-FM [14]	0.700	0.500	0.400	0.800	0.800	1.00	0.733
ORION-CMR (ViT-B8)	0.820	0.900	0.900	0.800	0.700	0.800	0.930
ViT-B16 (ablation)	0.740	0.600	0.500	0.900	0.900	0.800	0.870

Table 2. ACDC cine-SAX and EMIDEC LGE-SAX segmentation performance. Best results are shown in bold. (Dice (D), Hausdorff Distance (d_H), ViT-B = VB)

Method	ACDC cine-SAX						EMIDEC LGE-SAX			
	LV D	LV d_H	Myo D	Myo d_H	RV D	RV d_H	Myo D	Myo d_H	Scar D	Scar d_H
U-Net [23]	0.932	12.853	0.833	29.549	0.837	30.285	0.817	18.655	0.395	9.182
SoTA [2]	0.949	7.150	0.922	8.700	0.910	11.650	0.879	13.010	0.712	3.120
CMR-FM [14]	0.933	N/A	0.879	N/A	0.907	N/A	0.844	N/A	N/A	N/A
ORION-CMR (VB8)	0.942	8.527	0.878	16.417	0.899	13.606	0.915	6.110	0.799	3.034
VB16 (ablation)	0.921	9.897	0.834	21.561	0.877	13.829	0.818	9.153	0.653	7.550

Table 3. Subject-level sequence classification accuracy on the clinical cohort. SAX = short-axis; 2C/3C/4C = two-/three-/four-chamber.

Method	All	T2	T1	LGE	T2*	Loc.	Perf.	Flow	SAX	3C	2C	4C
ResNet[11]	0.925	0.033	0.456	0.761	0.850	0.867	0.878	0.896	0.975	1.000	0.974	0.986
ORION	0.992	0.975	0.991	0.990	1.000	0.999	0.998	0.998	0.993	1.000	0.973	0.979

3 Results

3.1 Public Benchmark Results

In all benchmarks, ViT-B8 consistently outperformed ViT-B16 (ablation) and was therefore selected for subsequent experiments. For LGE classification (Table 1), ORION-CMR achieved the highest performance on EMIDEC (0.930), surpassing both the previous SoTA (0.920) and CMR-FM (0.733). On ACDC, ORION-CMR exceeded CMR-FM (0.820 vs. 0.700) and significantly outperformed the supervised baseline (0.480, $P < 10^{-2}$).

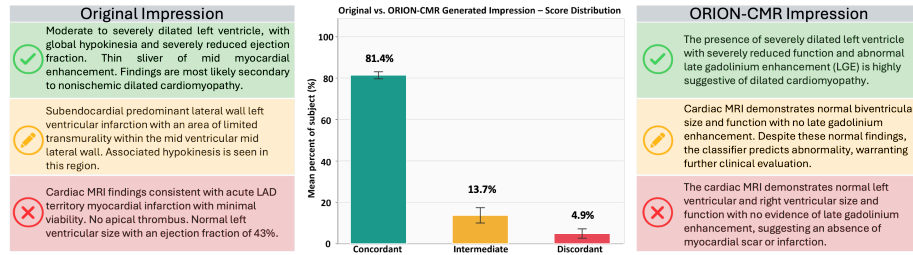


Fig. 2. Representative examples of original and ORION-CMR generated impressions, together with the distribution of review scores (N=68).

For segmentation (Table 2), ORION-CMR achieved D and d_H values comparable to the reported SoTA on ACDC, with differences (D , d_H) of LV (0.007, 1.377), myocardium (0.044, 7.717), and RV (0.011, 1.956), respectively. ORION-CMR established new SoTA performance, achieving Dice scores of 0.915 for myocardium and 0.799 for scar On EMIDEC, exceeding the previously reported SoTA (0.879 and 0.712, respectively). Compared with the supervised baseline, segmentation performance was significantly improved ($P < 10^{-2}$).

3.2 Report Evaluation

Representative examples of the original and generated reports, together with the radiologist review results, are shown in Fig. 2. Among the 68 generated reports, $81.4\% \pm 1.7\%$ were rated as concordant, $13.7\% \pm 3.7\%$ as intermediate, and $4.9\% \pm 2.2\%$ as discordant. Pairwise observed agreement ranged from 76.5% to 88.2%, with Cohen’s κ ranging from 0.41 to 0.61. Representative concordant and discordant cases are shown in Fig. 3, comparing the original clinical impression, automatically generated impression, and individual reader assessments. The complete scanner-native ORION-CMR pipeline, from image analysis through automated report generation, required around 90 seconds per subject. The complete scanner-native ORION-CMR pipeline, from image analysis through automated report generation, required around 90 seconds per subject. The average processing time was 11.2 ± 3.9 s for sequence classification, 51.6 ± 15.5 s for cine SAX segmentation, 9.7 ± 3.3 s for LGE and disease classifications and 11.2 ± 3.2 s for report generation.

Table 4. Clinical LGE and binary disease classification performance.

Metric	LGE Classification		Binary Disease Classification	
	ResNet-18 [11]	ORION-CMR	ResNet-18 [11]	ORION-CMR
AUC	0.706 (0.600–0.803)	0.826 (0.722–0.918)	0.761 (0.632–0.874)	0.960 (0.906–0.998)
Accuracy	0.650 (0.550–0.740)	0.797 (0.680–0.822)	0.765 (0.667–0.852)	0.912 (0.838–0.971)
Sensitivity	0.708 (0.660–0.885)	0.774 (0.622–0.913)	0.804 (0.691–0.906)	0.954 (0.881–1.000)
Specificity	0.520 (0.377–0.661)	0.816 (0.686–0.927)	0.680 (0.476–0.864)	0.840 (0.682–0.964)

Table 5. Per-class disease classification performance (one-vs-rest AUC).

Class	ResNet-18 [11]	ORION-CMR
Normal	0.761 (0.633–0.879)	0.960 (0.906–0.998)
Congenital heart disease	0.821 (0.730–0.913)	0.848 (0.667–1.000)
Dilated cardiomyopathy	0.674 (0.517–0.814)	0.845 (0.691–0.967)
Myocardial infarction	0.720 (0.594–0.836)	0.877 (0.767–0.964)

Original Clinical Impression	ORION-CMR Generated Impression	Reader Agreement
Concordant Examples		
Case 41. Normal cardiac chamber size and function (LVEF 56%) with no regional wall motion abnormality. No abnormal myocardial enhancement, edema, myocarditis, or infiltrative process. Increased LV apical trabeculation of uncertain clinical significance.	Case 41. Normal myocardial structure and function with no evidence of late gadolinium enhancement. In the absence of abnormal imaging findings and with low classifier probabilities for cardiomyopathy, the examination is most consistent with a normal heart.	Reader 1 Reader 2 Reader 3
Case 59. Findings consistent with non-ischemic dilated cardiomyopathy with severely reduced LV systolic function (LVEF 24%). No convincing LV myocardial delayed enhancement to suggest myocarditis.	Case 59. Severe LV dilation and severely reduced LV function are consistent with dilated cardiomyopathy. Abnormal myocardial enhancement is described as scar or fibrosis, although the clinical context does not strongly support acute myocardial infarction.	Reader 1 Reader 2 Reader 3
Case 39. Normal biventricular chamber size and global systolic function without regional wall motion abnormalities. No abnormal myocardial delayed enhancement. Similar nonspecific mediastinal lymphadenopathy.	Case 39. Normal myocardial structure and function with no evidence of late gadolinium enhancement. The absence of structural abnormalities and low probability of cardiomyopathy support classification as normal.	Reader 1 Reader 2 Reader 3
Discordant Examples		
Case 14. Acute LAD territory myocardial infarction with minimal viability. LVEF 43% with near-akinesis of the mid anteroseptal, mid anterior, and apical segments, corresponding myocardial edema and dense transmural scar, and mild microvascular obstruction.	Case 14. Normal LV and RV size and function with no evidence of LGE, suggesting absence of myocardial scar or infarction. The disease classifier nevertheless assigns a higher probability to myocardial infarction, creating an internal discrepancy.	Reader 1 Reader 2 Reader 3
Case 24. Type I L-R bicuspid aortic valve with mild-to-moderate eccentric aortic regurgitation. Mild LV dilation with preserved systolic function (LVEF 64%). No delayed enhancement to suggest ischemic or infiltrative cardiomyopathy. Normal RV size and function.	Case 24. Presence of LGE and mildly dilated ventricles is interpreted as supporting dilated cardiomyopathy, although congenital heart disease remains in the differential. Ventricular function is described as preserved despite structural abnormalities.	Reader 1 Reader 2 Reader 3
Case 33. No evidence of infiltrative cardiomyopathy, myocarditis, or myocardial scar. Moderately dilated LV with moderately depressed systolic function (LVEF 40%). Normal-sized RV with mildly depressed systolic function (RVEF 45%).	Case 33. Severe biventricular dilation with moderately reduced LV function is interpreted as dilated cardiomyopathy. LGE is described as indicating myocardial scar or fibrosis despite the original examination showing no myocardial scar.	Reader 1 Reader 2 Reader 3
Reader assessment: Concordant Intermediate Discordant		

Fig. 3. Representative original clinical impressions and ORION-CMR-generated impressions with independent expert-reader assessment. Three concordant and three discordant examples are shown. Green, orange, and red indicate concordant, intermediate, and discordant assessments, respectively.

4 Discussion

This study presents ORION-CMR, a self-supervised CMR foundation model, enabling MRI scanner-native end-to-end CMR interpretation. A single pretrained encoder supports multiple downstream tasks, including sequence classification, segmentation, LGE detection, disease classification, and automated report generation.

Across public benchmarks, ORION-CMR consistently outperformed supervised baselines and the previously published CMR-FM. The ViT-B8 encoder consistently outperformed ViT-B16, suggesting that finer patch representations better capture small anatomical structures and focal myocardial abnormalities. The complete workflow was successfully deployed on an MRI scanner using a lo-

cally hosted LLM, demonstrating the feasibility of real-time, privacy-preserving AI-assisted CMR interpretation.

Importantly, ORION-CMR was designed to reduce opportunities for unsupported LLM-generated findings. A shared pretrained ViT backbone performs the image-analysis tasks, with outputs consolidated into a deterministic structured representation. The rule-based LLM generates the final report solely from these predefined quantitative measurements and categorical findings, separating image interpretation from language generation.

This study has several limitations. Clinical validation was limited to a relatively small single-center cohort, particularly the Congenital heart disease subgroup ($n=7$), warranting further validation in larger, multi-institutional cohorts. Ventricular measurements were validated against values documented in the original clinical reports rather than expert-derived contours, which should be evaluated in future studies. Detailed ablation studies of individual pipeline components and further evaluation of the LLM prompting and structured-data schema are also warranted. The complete pipeline required approximately 90 seconds per subject on the scanner equipped with a GPU, supporting the feasibility of scanner-native deployment. Despite these limitations, to our knowledge, ORION-CMR represents the first fully automated, scanner-native CMR framework integrating image analysis, disease characterization, and report generation into a single end-to-end workflow.

In conclusion, ORION-CMR provides a unified framework for scanner-native image analysis and automated reporting, representing an important step toward clinically deployable end-to-end AI-assisted CMR interpretation.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Arnold, J.R., McCann, G.P.: Cardiovascular magnetic resonance: applications and practical considerations for the general cardiologist. *Heart* **106**(3), 174–181 (2020). <https://doi.org/10.1136/heartjnl-2019-314856>
2. Bernard, O., et al.: Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE Transactions on Medical Imaging* **37**(11), 2514–2525 (2018). <https://doi.org/10.1109/TMI.2018.2837502>
3. Bhatt, A.B., et al.: Congenital heart disease in the older adult: a scientific statement from the American Heart Association. *Circulation* **131**(21), 1884–1931 (2015). <https://doi.org/10.1161/CIR.0000000000000204>
4. Busch, F., Hoffmann, L., Dos Santos, D.P., et al.: Large language models for structured reporting in radiology: Past, present, and future. *European Radiology* (2025). <https://doi.org/10.1007/s00330-024-11107-6>
5. Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., Joulin, A.: Emerging properties in self-supervised vision transformers. In: *IEEE/CVF. pp. 9650–9660* (2021)

6. Catapano, F., et al.: Competence of radiologists in cardiac CT and MR imaging in europe: insights from the ESCR registry. *European Radiology* **34**(9), 5666–5677 (2024). <https://doi.org/10.1007/s00330-024-10644-4>
7. Demirel, O.B., Waddle, S., Ganji, S., Ooi, M., Robison, R.: Prime: An open innovation emitter–modulator–injector framework for inline mri reconstruction. In: *ISMRM* (2026)
8. Dweck, M.R., Williams, M.C., Moss, A.J., Newby, D.E., Fayad, Z.A.: Computed tomography and cardiac magnetic resonance in ischemic heart disease. *Journal of the American College of Cardiology* **68**(20), 2201–2216 (2016). <https://doi.org/10.1016/j.jacc.2016.08.047>
9. El Yaman, A., et al.: Temporal trends and geographic accessibility to cardiac magnetic resonance readers across the united states: an analysis of Medicare Part B data. *Journal of Cardiovascular Magnetic Resonance* **27**(2), 101921 (2025). <https://doi.org/10.1016/j.jocmr.2025.101921>
10. Hamamci, I.E., et al.: Generalist foundation models from a multimodal dataset for 3D computed tomography. *Nature Biomedical Engineering* (2026). <https://doi.org/10.1038/s41551-025-01599-y>
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 770–778 (2016)
12. Heidenreich, P.A., et al.: 2022 AHA/ACC/HFSA guideline for the management of heart failure. *Journal of the American College of Cardiology* **79**(17), e263–e421 (2022). <https://doi.org/10.1016/j.jacc.2021.12.012>
13. Hundley, W.G., et al.: ACCF/ACR/AHA/NASCI/SCMR 2010 expert consensus document on cardiovascular magnetic resonance. *Journal of the American College of Cardiology* **55**(23), 2614–2662 (2010). <https://doi.org/10.1016/j.jacc.2009.11.011>
14. Jacob, A.J., Borgohain, I., Chitiboi, T., Sharma, P., Comaniciu, D., Rueckert, D.: Towards a cardiovascular magnetic resonance foundation model for multi-task cardiac image analysis. *Journal of Cardiovascular Magnetic Resonance* **27**(2), 101967 (2025). <https://doi.org/10.1016/j.jocmr.2025.101967>
15. Kawel-Boehm, N., Hetzel, S.J., Ambale-Venkatesh, B., et al.: Society for cardiovascular magnetic resonance reference values ("normal values") in cardiovascular magnetic resonance: 2025 update. *Journal of Cardiovascular Magnetic Resonance* **27**(1), 101853 (2025). <https://doi.org/10.1016/j.jocmr.2025.101853>
16. Lalande, A., et al.: EMIDEC: a database usable for the automatic evaluation of myocardial infarction from delayed-enhancement cardiac MRI. *Data* **5**(4), 89 (2020). <https://doi.org/10.3390/data5040089>
17. Lalande, A., Chen, Z., Pommier, T., Decourselle, T., Qayyum, A., Salomon, M., Ginjac, D., Skandarani, Y., Boucher, A., Brahim, K., et al.: Deep learning methods for automatic evaluation of delayed enhancement-mri. the results of the emidec challenge. *Medical Image Analysis* **79**, 102428 (2022)
18. Meyer, A., Murali, A., Zarin, F., Mutter, D., Padoy, N.: UltraSAM: a foundation model for ultrasound using large open-access segmentation datasets. *International Journal of Computer Assisted Radiology and Surgery* **21**(1), 93–102 (2025). <https://doi.org/10.1007/s11548-025-03517-8>
19. Natale, L., et al.: Cardiac radiology in europe: status and vision by the European Society of Cardiovascular Radiology (ESCR) and the European Society of Radiology (ESR). *European Radiology* **33**(8), 5489–5497 (2023). <https://doi.org/10.1007/s00330-023-09533-z>

20. Oquab, M., et al.: DINOv2: learning robust visual features without supervision. arXiv preprint arXiv:2304.07193 (2024). <https://doi.org/10.48550/arXiv.2304.07193>
21. Raisi-Estabragh, Z., Szabo, L., McCracken, C., Bülow, R., Aquaro, G.D., Andre, F., Le, T.T., Suchá, D., Condurache, D.G., Salih, A.M., Chadavada, S., Aung, N., Lee, A.M., Harvey, N.C., Leiner, T., Chin, C.W.L., Friedrich, M.G., Barison, A., Dörr, M., Petersen, S.E.: Cardiovascular magnetic resonance reference ranges from the healthy hearts consortium. *JACC: Cardiovascular Imaging* **17**(7), 746–762 (2024). <https://doi.org/10.1016/j.jcmg.2024.01.009>
22. Rajiah, P.S., François, C.J., Leiner, T.: Cardiac MRI: state of the art. *Radiology* **307**(3), e223008 (2023). <https://doi.org/10.1148/radiol.223008>
23. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
24. Ruijsink, B., et al.: Fully automated, quality-controlled cardiac analysis from cardiovascular magnetic resonance using deep learning. *Journal of Cardiovascular Magnetic Resonance* **22**(1), 65 (2020). <https://doi.org/10.1186/s12968-020-00656-8>
25. Savage, C.H., Kanhere, A., Parekh, V., et al.: Open-source large language models in radiology: A review and tutorial for practical research and clinical deployment. *Radiology* (2025)
26. Wang, Y., et al.: Artificial intelligence for automated screening and diagnosis of cardiovascular disease using cine and late gadolinium enhancement cardiac mri. *Radiology: Artificial Intelligence* (2024)
27. Woźnicki, P., Laqua, C., Fiku, I., et al.: Automatic structuring of radiology reports with on-premise open-source large language models. *European Radiology* (2025)
28. Zhan, Y., et al.: Normal reference values for cardiac magnetic resonance ventricular volumes, function, and mass: A systematic review and meta-analysis. *European Radiology* (2024), meta-analysis of 12,812 healthy adults from 25 countries
29. Zhang, Y., et al.: Towards cardiac MRI foundation models: comprehensive visual-tabular representations for whole-heart assessment and beyond. *Medical Image Analysis* **106**, 103756 (2025). <https://doi.org/10.1016/j.media.2025.103756>